

---

# Knowledge Management and Taxonomies

Dr. Geoffrey P Malafsky, TechI<sup>2</sup> LLC

# Context: Who Are We?

---

- Technology Architect: Knowledge and Content Management Systems
  - Dept of Navy CIO: Taxonomy, systems, business processes, metrics
  - Lifelines Portal: redesign of 20 million hits/month public site to Oracle architecture with workflow + portal + taxonomy + search
- Independent Assessments and Planning
  - Balanced Scorecards: technology ↔ resources ↔ processes
  - External Independent Review: in-depth objective analysis of all good and bad aspects of system or plan
  - R&D Expert:
    - DARPA: LCCMD (MEMS), SUO-SAS (mobile peer-to-peer data fusion and dissemination); CPOF (KM and visualization)
    - ONR: Swampworks; CINC21

# The Need

---

- Share and reuse knowledge
- Reduce information overload
- Minimize operation and maintenance costs
- Streamline business processes
- Enterprise system interoperability

# The Challenge

---

- Enormous quantity of written, spoken, and visual information
- Confusion about what “knowledge” is
- Language complexity and dynamism
- Limited tool accuracy
- Multiple systems and legacy applications

# An Approach

---

- Develop and implement the proper set of taxonomies for the organization, workflow, and subject matter
- Incorporate contextual meaning into system architecture with business logic

# What is Knowledge?

---

- Context: what is it about?
- Confidence: is it right?
- Relationships: what does it have to do with that?
- Priorities: what is most important?
- Types
  - Explicit knowledge is codified and can be manipulated
  - Tacit knowledge is unspoken “know-how”

# Knowledge Ontology:Bloom 1956

---

- Knowledge
  - Knowledge of specifics
    - knowledge of terminology
    - knowledge of specific facts
  - Knowledge of ways and means of dealing with specifics
    - knowledge of conventions
    - knowledge of trends and sequences
    - knowledge of classifications and categories

# Knowledge is Personal

---

- Depends on prior knowledge, and task focus
- “Set the soldering iron to 200 degrees”
  - *information* from manual for general use
  - *knowledge* from expert for specific manufacturing process
- “10000 units shipped yesterday”
  - *data* for logistics
  - *information* for shipping manager
  - *knowledge* for competitor monitoring market share



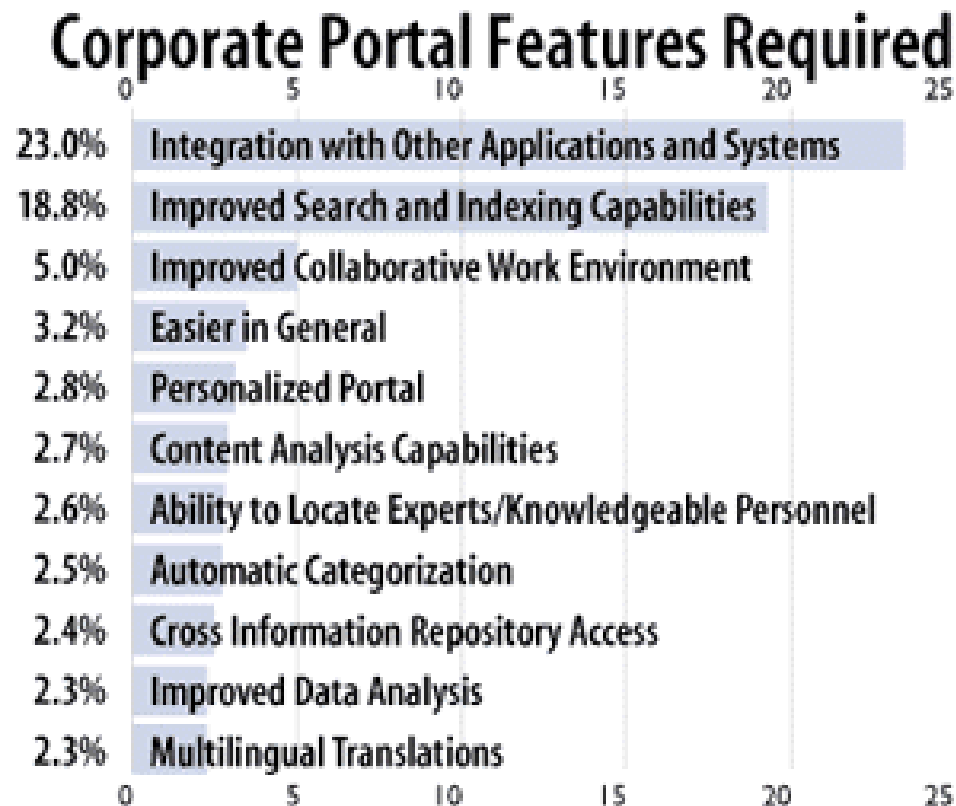
# Language Complexity

---



- People need the metamessages conveyed by human interactions
  - body language
  - clothing
  - dialect
  - intonation

# Where is it?



Source: Delphi Group 1999

# Semantic Web

---

- The vision of the *Semantic Web* aims at creating a Web where information can be “understood” by machines as well as humans
- The *Semantic Web* requires the emergence of a general purpose representation and markup-language to convey information about machine-accessible semantics
- Taxonomies can be expressed as abbreviated Ontologies for the *Semantic Web*

# Potential KM System Solutions

---

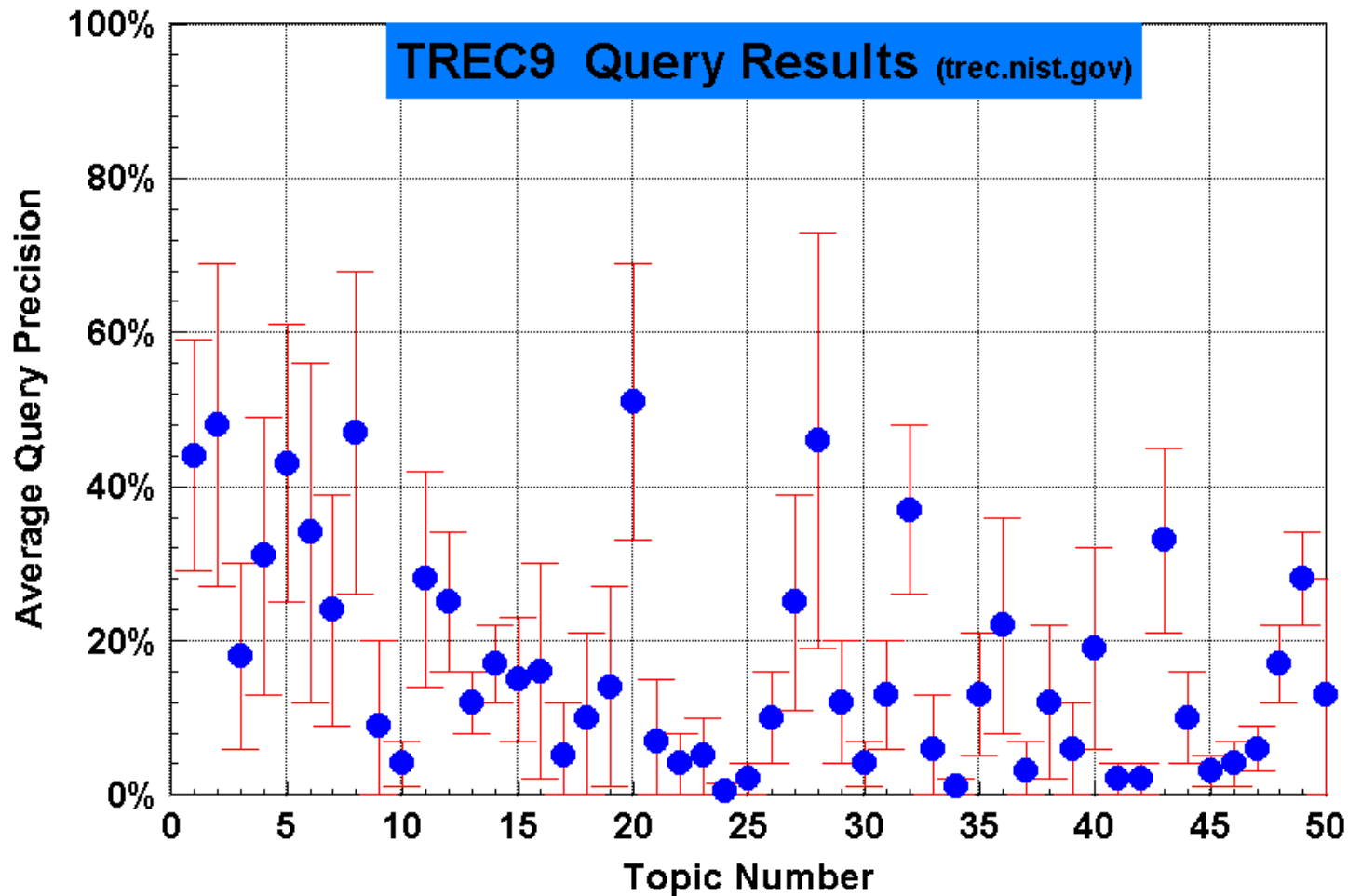
- Use people: librarians, intelligence analysts
  - Very effective but costly and need time
  - Consistency and interoperability
- Limit content: small validated set
  - Hard to maintain currency
  - Large effort to collect, review, organize
  - Who decides what is authoritative?

# Potential KM System Solutions

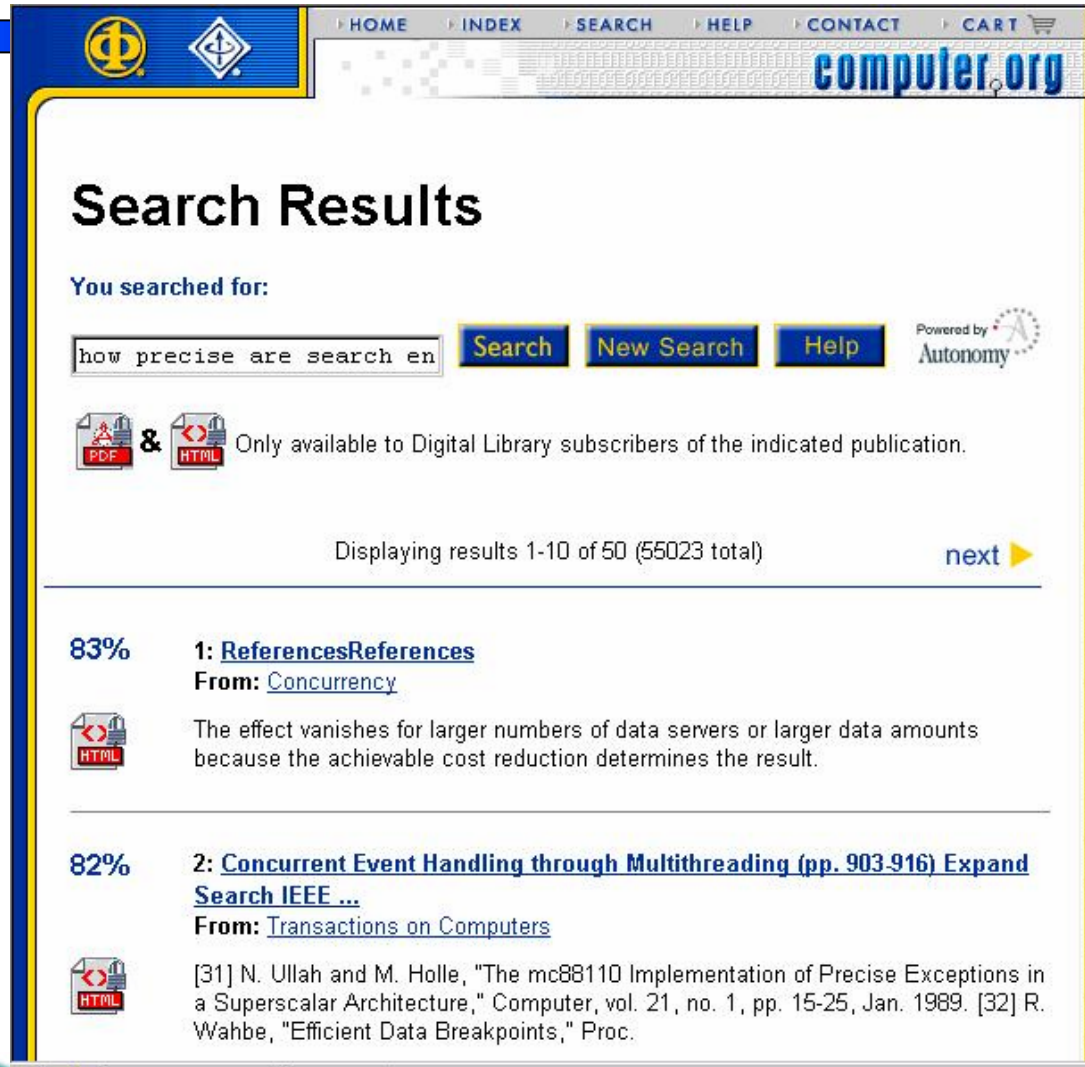
---

- Technology: search (unstructured text)
  - Low maintenance effort but low precision
  - Depends on scaling and information density
- Use categories
  - creates smaller sets and introduces structure
    - ancient technique: Pinakes' Callimachus
  - **TAXONOMY defines topics and their relationships**
    - **improves user and technology effectiveness**

# Search Engines: State-of-the-Art



# Tool Performance



The screenshot shows a web browser window displaying search results on the computer.org website. The page has a blue header with navigation links: HOME, INDEX, SEARCH, HELP, CONTACT, and CART. The search results section is titled "Search Results" and shows the search query "how precise are search en". Below the query are buttons for "Search", "New Search", and "Help". A "Powered by Autonomy" logo is visible. The results are displayed in a list format, with the first result showing a PDF icon and the text "Only available to Digital Library subscribers of the indicated publication." The second result is titled "1: ReferencesReferences" and "From: Concurrency". The third result is titled "2: Concurrent Event Handling through Multithreading (pp. 903-916) Expand Search IEEE ..." and "From: Transactions on Computers". The page also shows a "next" button and a "Displaying results 1-10 of 50 (55023 total)" message.



HOME INDEX SEARCH HELP CONTACT CART

computer.org

## Search Results


You searched for:

how precise are search en [Search](#) [New Search](#) [Help](#) Powered by Autonomy


  Only available to Digital Library subscribers of the indicated publication.

Displaying results 1-10 of 50 (55023 total) [next](#)

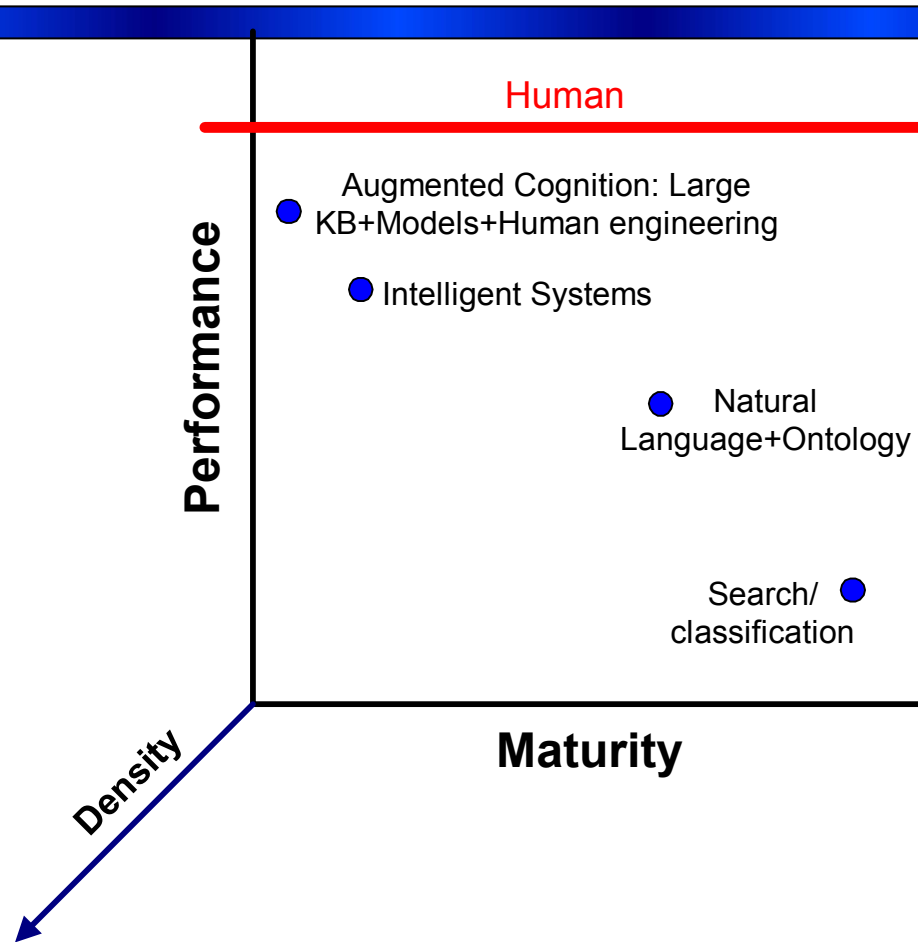
**83%** **1: [ReferencesReferences](#)**  
**From: [Concurrency](#)**

 The effect vanishes for larger numbers of data servers or larger data amounts because the achievable cost reduction determines the result.

**82%** **2: [Concurrent Event Handling through Multithreading \(pp. 903-916\) Expand Search IEEE ...](#)**  
**From: [Transactions on Computers](#)**

 [31] N. Ullah and M. Holle, "The mc88110 Implementation of Precise Exceptions in a Superscalar Architecture," Computer, vol. 21, no. 1, pp. 15-25, Jan. 1989. [32] R. Wahbe, "Efficient Data Breakpoints," Proc.

# Technology Performance Capabilities





# Organizing Information

---

- Ontologies and taxonomies
  - concepts and descriptions
- Develop enterprise architecture for organizational taxonomies
  - Every workgroup naturally develops its own most efficient schema
- People mentally organize in multiple ways based on task and interest

# Taxonomy Definition: APQC

---

- ✓ **A classification scheme for the knowledge accessible through a given system or interface (ultimately multi-dimensional)**
- ✓ **Facilitates effective retrieval, capturing, and recognition of content that is important to target users**
- ✓ **A taxonomy typically includes:**
  - ✓ **A navigable hierarchy of concepts and terms**
  - ✓ **Information “tags” that further identify and categorize content elements**
- ✓ **Links from the taxonomy lead to resources (e.g., people, documents, and events)**
  - ✓ **May or may not also include a thesaurus**

# Datawarehouse MetaData

## Examples of Technical Meta Data

User report and query access patterns, frequency, and execution time

System audit controls and balancing information

The system of record feeding the data warehouse

Identification of source system fields

Mappings and transformations from the system of record to the data warehouse

## Examples of Business Meta Data

The structure of data as known to the business analyst

Common access routines for information in the warehouse/mart

Table names and definitions in business terms


Attribute names and definitions in business terms

*David Marco, Enterprise Warehousing Solutions, Inc., January 1999*

# Taxonomy Complexity

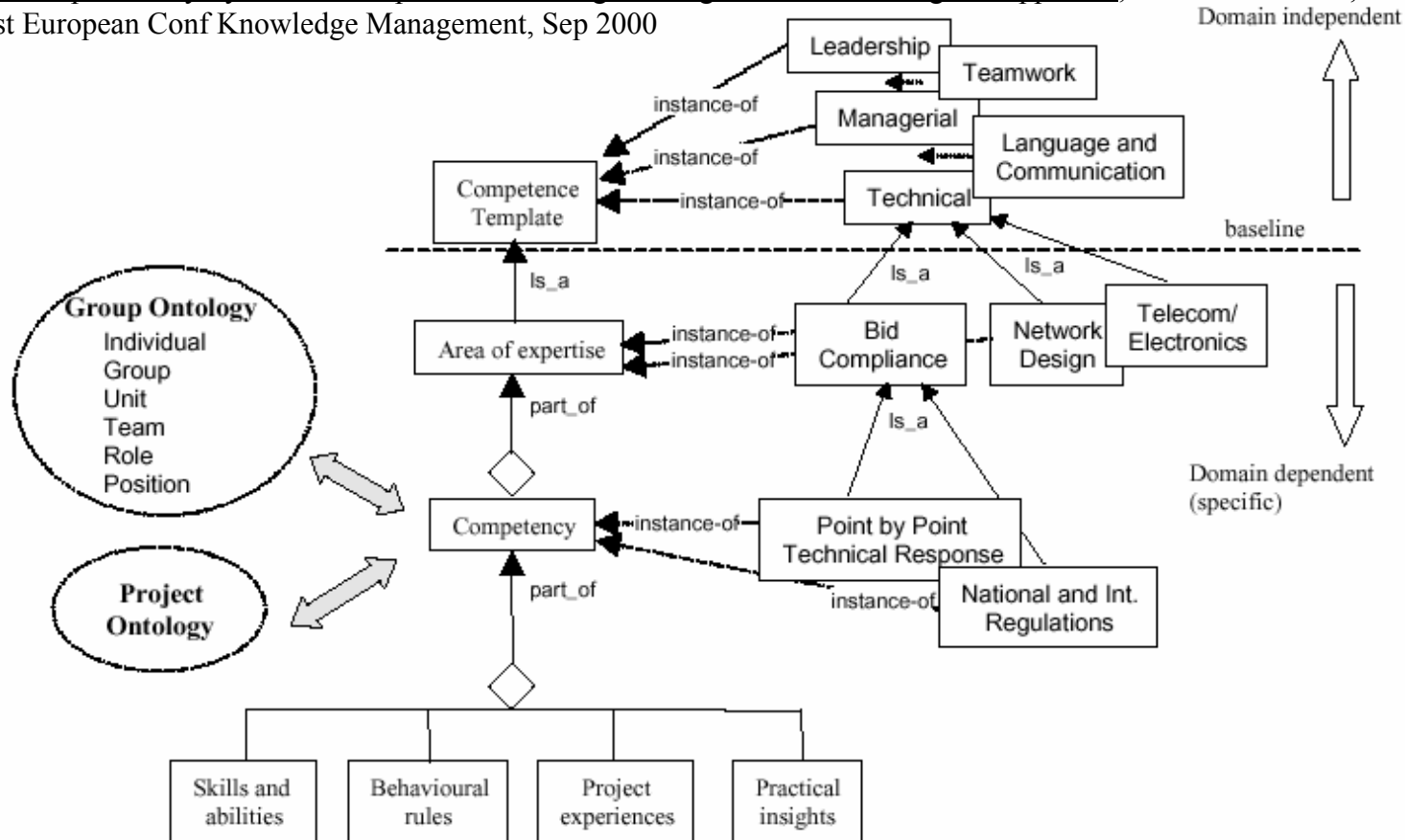
## 80. INTERDISCIPLINARY PHYSICS AND RELATED AREAS OF SCIENCE AND TECHNOLOGY

### 81. Materials science

- 81.05.  Specific materials: fabrication, treatment, testing and analysis
  - ▽▽▽▽/ *Superconducting materials, see 74.70 and 74.72*
  - ▽▽▽▽/ *Magnetic materials, see 75.50*
  - ▽▽▽▽/ *Optical materials, see 42.70*
  - ▽▽▽▽/ *Dielectric, piezoelectric, and ferroelectric materials, see 77.80*
  - ▽▽▽▽/ *Colloids, gels, and emulsions, see 82.70.D, G, K respectively*
  - ▽▽▽▽/ *Biological materials, see 87.14*
- 81.05.Bx Metals, semimetals, and alloys
- 81.05.Cy Elemental semiconductors
- 81.05.Dz II–VI semiconductors
- 81.05.Ea III–V semiconductors
- 81.05.Gc Amorphous semiconductors
- 81.05.Hd Other semiconductors
- 81.05.Je Ceramics and refractories (including borides, carbides, hydrides, nitrides, oxides, and silicides)
- 81.05.Kf Glasses (including metallic glasses)
- 81.05.Lg Polymers and plastics; rubber; synthetic and natural fibers; organometallic and organic materials
- 81.05.Mh Cermets, ceramic and refractory composites
- 81.05.Ni Dispersion-, fiber-, and platelet-reinforced metal-based composites

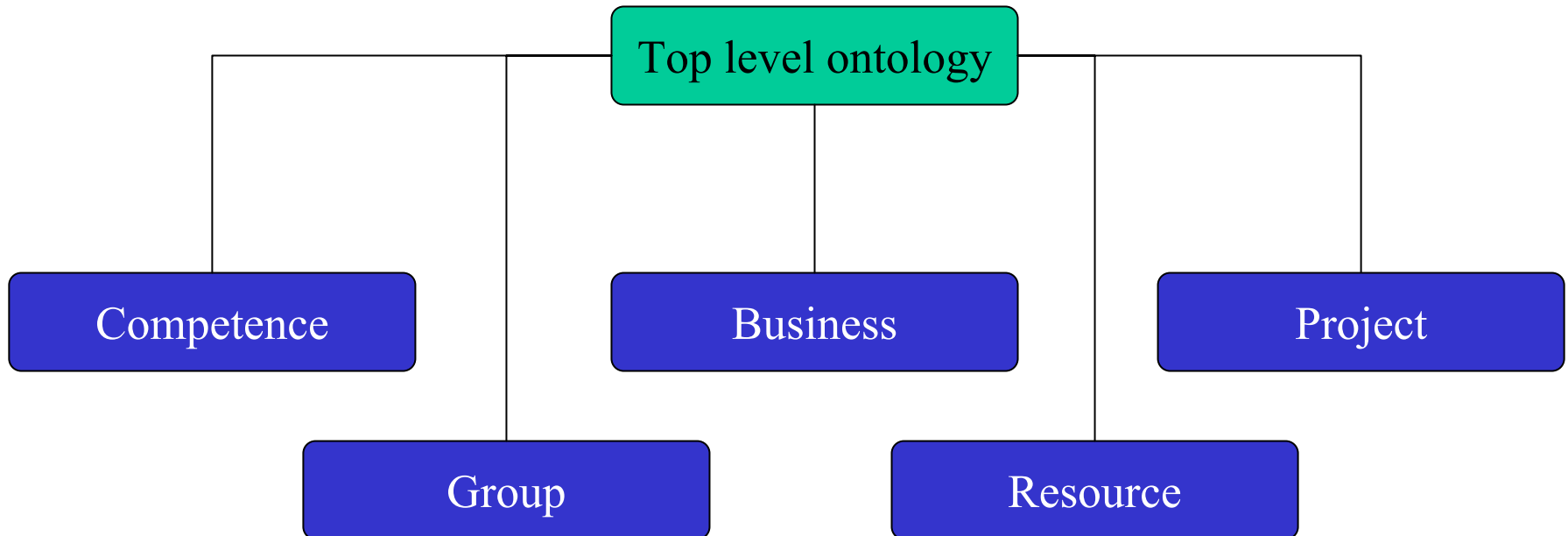
# Organization Requires Context

A Group Memory System for Corporate Knowledge Management: An Ontological Approach, José Vasconcelos, et al, Proc 1st European Conf Knowledge Management, Sep 2000

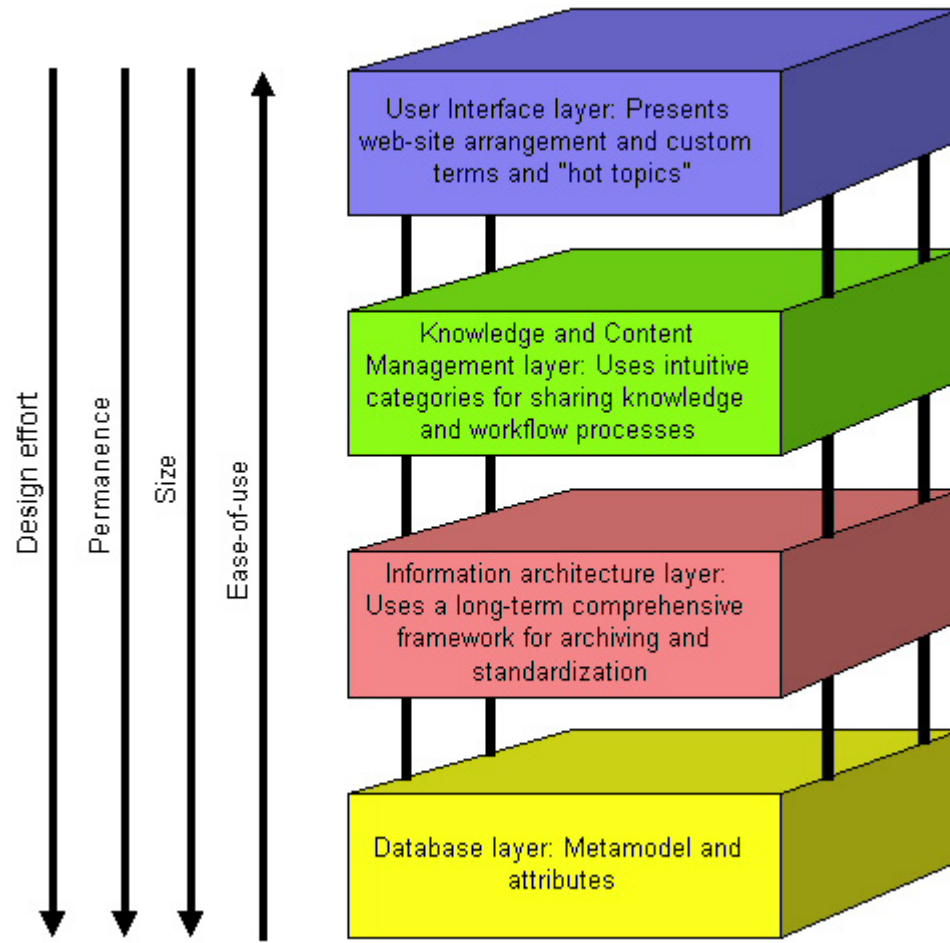


# Different Perspectives

A Group Memory System for Corporate Knowledge Management: An Ontological Approach, José Vasconcelos, et al, September 2000,



# Types of Taxonomies



# KM Taxonomy

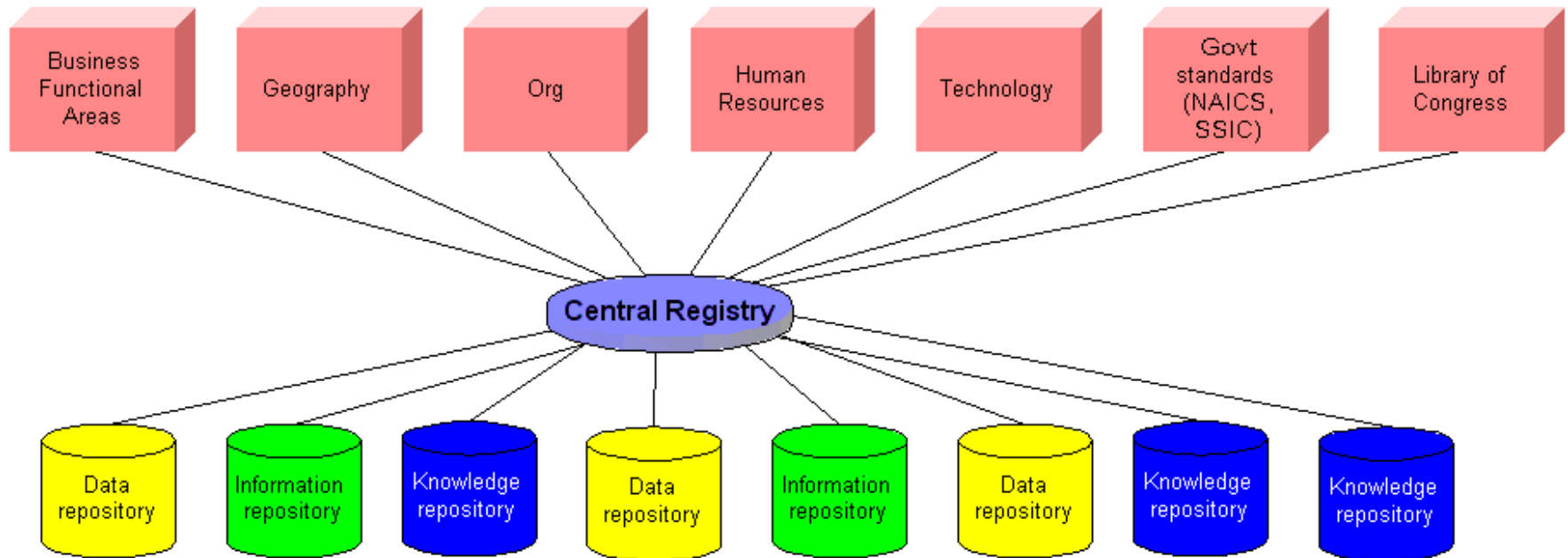
---

## **What is a KM taxonomy?**

A taxonomy is a structured set of names and descriptions used to organize sources in a consistent way. A typical taxonomy uses a logical arrangement but doesn't account for users' particular decision-making and action-taking needs. A KM taxonomy focuses on enabling efficient and interoperable retrieval and sharing of data, information, and knowledge across the enterprise by building in natural workflow and knowledge needs in an intuitive structure.



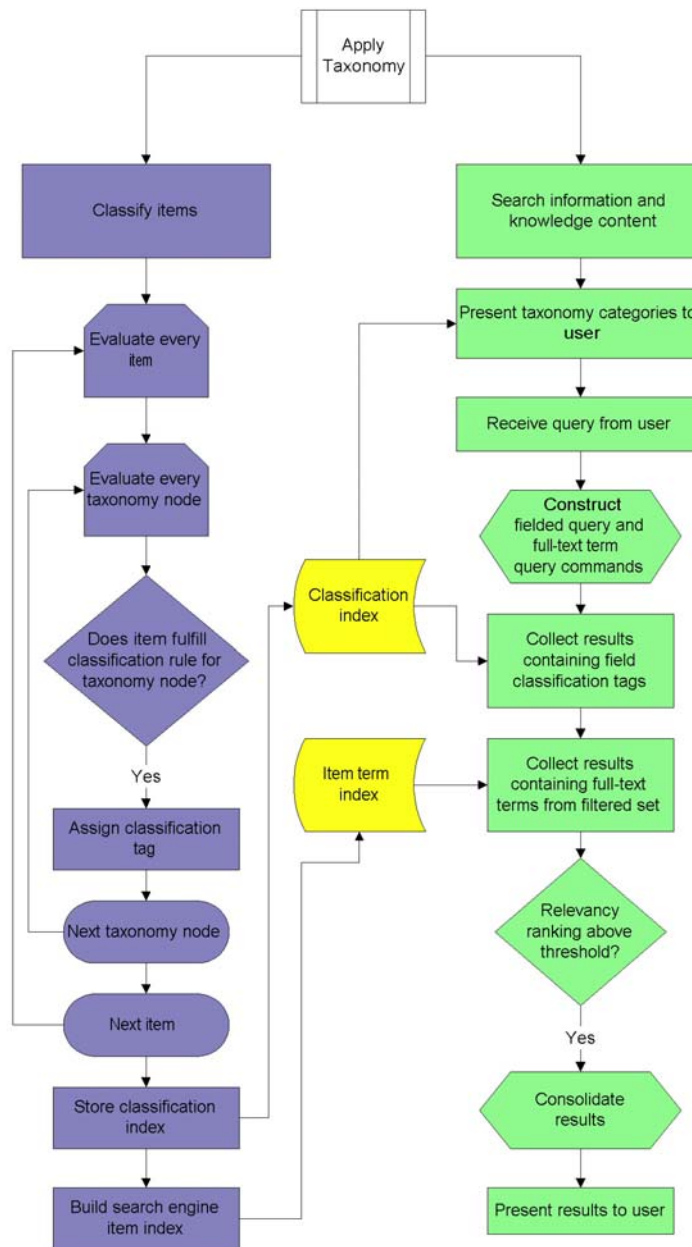
# Multiple Domains



# Applying Taxonomy

---

- Must associate taxonomy nodes with every document/record/multimedia/...
- How accurate is the assignment?
- Taxonomy use: What/why/who/
  - work processes and applications
- Where: HTML header, metadata repository
- How:
  - Fielded search engine query
  - Middleware



# Example

---

**Lessons Learned**  
*The Information Source*

**In Body:**

<b>SAIC lead</b>	<input type="text" value="all"/>	<input type="button" value="must include"/>	<b>Common Questions</b>	<input type="text" value="Choose a commonly asked question"/>
<b>Customer</b>	<input type="text"/>	<input type="button" value="must include"/>	<b>SAICFY</b>	<input type="text" value="any"/> <input type="button" value="must include"/>
<b>Winner</b>	<input type="text"/>	<input type="button" value="must include"/>	<b>GOVIFY</b>	<input type="text" value="any"/> <input type="button" value="must include"/>
<b>Incumbent</b>	<input type="text"/>	<input type="button" value="must include"/>	<b>Cal Yr</b>	<input type="text" value="any"/> <input type="button" value="must include"/>
<b>Competitor</b>	<input type="text"/>	<input type="button" value="must include"/>	<b>Won/lost</b>	<input type="text" value="either"/> <input type="button" value="must include"/>
<b>In LL</b>	<input type="text"/>	<input type="button" value="must include"/>	<b>Contract</b>	<input type="text" value="any"/> <input type="button" value="must include"/>
<b>Major LL</b>	<input type="text" value="any"/>	<input type="button" value="must include"/>	<b>Notes:</b> Use multiple filters by entering or choosing search terms. These are restricted to the respective fields. For exact phrases, use quotations ("past performance"). <u>Common Questions</u> supercedes all other query filters.	
<b>Topics</b>	<input type="text"/>	<input type="button" value="must include"/>		

# Demonstration: Typical Search

**Lessons Learned** The Information Source

**In Body:**

SAIC lead	<input type="text" value="all"/>	<input type="button" value="must include"/>	<b>Common Questions</b>	<input type="text" value="Choose a commonly asked question"/>
Customer	<input type="text"/>	<input type="button" value="must include"/>	SAICFY	<input type="text" value="any"/> <input type="button" value="must include"/>
Winner	<input type="text"/>	<input type="button" value="must include"/>	GOVIFY	<input type="text" value="any"/> <input type="button" value="must include"/>
Incumbent	<input type="text"/>	<input type="button" value="must include"/>	Cal Yr	<input type="text" value="any"/> <input type="button" value="must include"/>
Competitor	<input type="text"/>	<input type="button" value="must include"/>	Won/lost	<input type="text" value="either"/> <input type="button" value="must include"/>
In LL	<input type="text"/>	<input type="button" value="must include"/>	Contract	<input type="text" value="any"/> <input type="button" value="must include"/>
Major LL	<input type="text" value="any"/>	<input type="button" value="must include"/>	<b>Notes:</b> Use multiple filters by entering or choosing search terms. These are restricted to the respective fields. For exact phrases, use quotations ("past performance"). <u>Common Questions</u> supercedes all other query filters.	
Topics	<input type="text"/>	<input type="button" value="must include"/>		

Results for: poor past performance

637/1000 returned

Document count: poor (72) past (554) performance (600) poor past... (9)


637 results found, top 500 sorted by relevance

[sort by date](#) [hide summaries](#)

1-25

## High Performance Computing Center (sub sa 5, M) [Fiscal Year: 2001 - Period 13 Loss]

The purpose of this contract is to provide High Performance Computing Center architecture, design, development, and operation for ...  
file:///d:/current projects/lldb/lldb collection/High Performance Computing Cen-Cat.htm - 31.8KB - poor: 2, past: 5, performance: 16, poor past...: 5

70%   
05 Apr 01

## WSMR Analysis and Engineering Support [Fiscal Year: 1998 - Period 3 Loss]

The overall evaluation results are as shown below. The evaluation was a numeric

file:///d:/current projects/lldb/lldb collection/WSMR Analysis and Engineering Cat.htm - 22.8KB - poor: 2, past: 5, performance: 16, poor past...: 5

68%   
05 Jan 00

# Demonstration: 1 KM Category

## Lessons Learned

The Information Source

In Body:

Search

Reset

SAIC lead

Common Questions

Customer

SAICFY

Winner

GOVIFY

Incumbent

Cal Yr

Competitor

Won/lost

In LL

Contract

Major LL

Topics

Notes: Use multiple filters by entering or choosing search terms. These are restricted to the respective fields. For exact phrases, use quotations ("past performance"). Common Questions supercedes all other query filters.

Results for: +majorll:"past performance"

95/1000 returned

95 results found, sorted by relevance

[sort by date](#) [hide summaries](#)

1-25

### World-Wide Information Technology Support (Answer) (ma) [Fiscal Year: 1999 - Period 13 Win]

36%

The intent of these contracts is to provide federal agencies a variety of Information Technology support services. These ...

03 Jan 00

file:///d:/current projects/lldb/lldb collection/World-Wide Information Technol-Cat.htm - 42.1KB - +majorll:"past performance": 2

### WSMR Analysis and Engineering Support [Fiscal Year: 1998 - Period 3 Loss]

36%

The overall evaluation results are as shown below. The evaluation was a numeric

05 Jan 00



# Demonstration: 2 Intuitive Categories

**Lessons Learned** The Information Source

In Body:

SAIC lead	all	must include	Common Questions	Choose a commonly asked question	
Customer	<input type="text"/>	must include	SAICFY	any	must include
Winner	<input type="text"/>	must include	GOVIFY	any	must include
Incumbent	<input type="text"/>	must include	Cal Yr	any	must include
Competitor	<input type="text"/>	must include	Won/lost	won	must include
In LL	<input type="text"/>	must include	Contract	any	must include
Major LL	Past performance	must include	<b>Notes:</b> Use multiple filters by entering or choosing search terms. These are restricted to the respective fields. For exact phrases, use quotations ("past performance"). <u>Common Questions</u> supercedes all other query filters.		
Topics	<input type="text"/>	must include			

Results for: +majorll:"past performance" +wonloss:win

44/1000 returned

ll:"past performance" (95) +wonloss:win (516)

44 results found, sorted by relevance

[sort by date](#) [hide summaries](#)

1-25

## World-Wide Information Technology Support (Answer) (ma) [Fiscal Year: 1999 - Period 13 Win]

38%

The intent of these contracts is to provide federal agencies a variety of Information Technology support services. These ...

03 Jan 00

file:///d:/current projects/lldb/lldb collection/World-Wide Information Technol-Cat.htm - 42.1KB - +majorll:"past performance": 2, +wonloss:win: 2

## Total Environmental Restoration Contract (sa) [Fiscal Year: 2000 - Period 4 Win]

38%

Montgomery Watson was awarded a \$300M, ten year contract in April 1997 to provide the full range of studies, design, construction, ...

14 Oct 99

file:///d:/current projects/lldb/lldb collection/Total Environmental Restoratio-Cat.htm - 24.5KB - +majorll:"past performance": 2, +wonloss:win: 2

# Demonstration: Simply Add More

**Lessons Learned** The Information Source

In Body:

SAIC lead	all	must include	Common Questions	Choose a commonly asked question
Customer	<input type="text"/>	must include	SAICFY	any
Winner	<input type="text"/>	must include	GOVTFY	2001
Incumbent	<input type="text"/>	must include	Cal Yr	any
Competitor	<input type="text"/>	must include	Won/lost	won
In LL	<input type="text"/>	must include	Contract	any
Major LL	Past performance	must include	<b>Notes:</b> Use multiple filters by entering or choosing search terms. These are restricted to the respective fields. For exact phrases, use quotations ("past performance"). <u>Common Questions</u> supercedes all other query filters.	
Topics	<input type="text"/>	must include		

Results for: +majorll:"past performance" +govtfy:2001 +wonloss:win

3/1000 returned !!

3 results found, sorted by relevance

[sort by date](#) [hide summaries](#)

1-3

**Supp of Dev and Analysis of Specific Contracts (ma 12, D) [Fiscal Year: 2002 - Period 1 Win]**

43%

The purpose of these contracts is to give the Joint Experimentation Directorate the ability to reach out and get

08 May 01

file:///d:/current projects/lldb/lldb collection/Supp of Dev and Analysis of Sp-Cat.htm - 28.8KB - +majorll:"past performance": 2, +govtfy:2001: 1, +wonloss:win: 2

**Stricom Omnibus Contract - Virtual Lot (ma 10, P) [Fiscal Year: 2002 - Period 1 Win]**

43%

SAIC did a lot of things right in pursuing these awards. These included a 3rd Party Client Assessment, Pre-Proposal Workshop, a ...

05 May 01



# Conclusions

---

- Automated search/categorization is still not precise enough
- Taxonomies allow large information quantity to be separated into more manageable chunks
- There are different types of taxonomies
- Effectiveness and usability require KM Taxonomy of intuitive categories